

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.

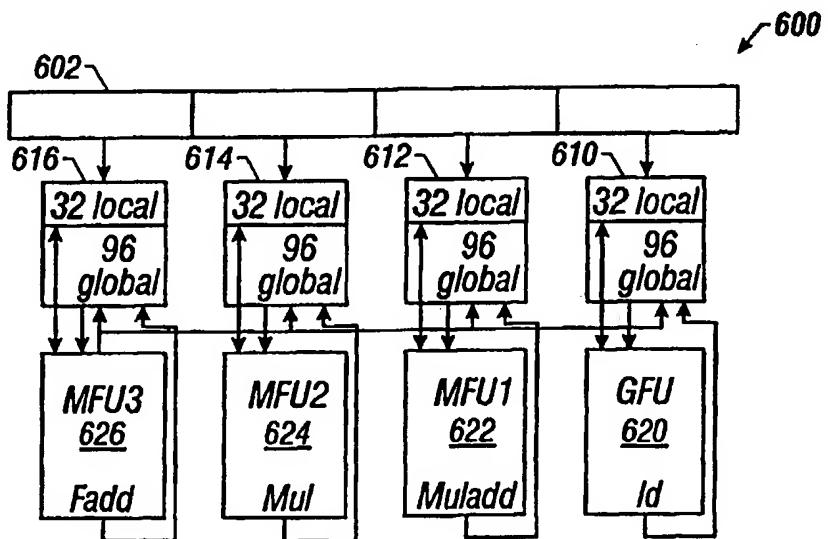
PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

| | | |
|---|---|--|
| (51) International Patent Classification ⁷ : G06F 9/30 | A1 | (11) International Publication Number: WO 00/33178 (43) International Publication Date: 8 June 2000 (08.06.00) |
| (21) International Application Number: PCT/US99/28820 (22) International Filing Date: 3 December 1999 (03.12.99) (30) Priority Data: 09/204,585 3 December 1998 (03.12.98) US | (81) Designated States: JP, KR, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE). Published <i>With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i> | |
| (71) Applicant: SUN MICROSYSTEMS, INC. [US/US]; 901 San Antonio Road, M/S PAL01-521, Palo Alto, CA 94303 (US). (72) Inventors: TREMBLAY, Marc; 140 Hanna Way, Menlo Park, CA 94025 (US). JOY, William, N.; P.O. Box 23, Aspen, CO 81612-0023 (US). (74) Agents: KOESTNER, Ken, J. et al.; Skjerven, Morrill, MacPherson, Franklin & Friel LLP, Suite 700, 25 Metro Drive, San Jose, CA 95110 (US). | | |
| (54) Title: LOCAL AND GLOBAL REGISTER PARTITIONING IN A VLIW PROCESSOR | | |
| <p>(57) Abstract</p> <p>A Very Long Instruction Word (VLIW) processor having a plurality of functional units includes a multiported register file that is divided into a plurality of separate register file segments, each of the register file segments being associated to one of the plurality of functional units. The register file segments are partitioned into local registers and global registers. The global registers are read and written by all functional units. The local registers are read and written only by a functional unit associated with a particular register file segment. The local registers and global registers are addressed using register addresses in an address space that is separately defined for a register file segment/functional unit pair. The global registers are addressed within a selected global register range using the same register addresses for the plurality of register file segment/functional unit pairs. The local registers in a register file segment are addressed using register addresses in a local register range outside the global register range that are assigned within a single register file segment/functional unit pair. Register addresses in the local register range are the same for the plurality of register file segment/functional unit pairs and address registers locally within a register file segment/functional unit pair.</p> | | |



FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

| | | | | | | | |
|----|--------------------------|----|--|----|--|----|--------------------------|
| AL | Albania | ES | Spain | LS | Lesotho | SI | Slovenia |
| AM | Armenia | FI | Finland | LT | Lithuania | SK | Slovakia |
| AT | Austria | FR | France | LU | Luxembourg | SN | Senegal |
| AU | Australia | GA | Gabon | LV | Latvia | SZ | Swaziland |
| AZ | Azerbaijan | GB | United Kingdom | MC | Monaco | TD | Chad |
| BA | Bosnia and Herzegovina | GE | Georgia | MD | Republic of Moldova | TG | Togo |
| BB | Barbados | GH | Ghana | MG | Madagascar | TJ | Tajikistan |
| BE | Belgium | GN | Guinea | MK | The former Yugoslav Republic of Macedonia | TM | Turkmenistan |
| BF | Burkina Faso | GR | Greece | ML | Mali | TR | Turkey |
| BG | Bulgaria | HU | Hungary | MN | Mongolia | TT | Trinidad and Tobago |
| BJ | Benin | IE | Ireland | MR | Mauritania | UA | Ukraine |
| BR | Brazil | IL | Israel | MW | Malawi | UG | Uganda |
| BY | Belarus | IS | Iceland | MX | Mexico | US | United States of America |
| CA | Canada | IT | Italy | NE | Niger | UZ | Uzbekistan |
| CF | Central African Republic | JP | Japan | NL | Netherlands | VN | Viet Nam |
| CG | Congo | KE | Kenya | NO | Norway | YU | Yugoslavia |
| CH | Switzerland | KG | Kyrgyzstan | NZ | New Zealand | ZW | Zimbabwe |
| CI | Côte d'Ivoire | KP | Democratic People's Republic of Korea | PL | Poland | | |
| CM | Cameroon | KR | Republic of Korea | PT | Portugal | | |
| CN | China | KZ | Kazakhstan | RO | Romania | | |
| CU | Cuba | LC | Saint Lucia | RU | Russian Federation | | |
| CZ | Czech Republic | LI | Liechtenstein | SD | Sudan | | |
| DE | Germany | LK | Sri Lanka | SE | Sweden | | |
| DK | Denmark | LR | Liberia | SG | Singapore | | |
| EE | Estonia | | | | | | |

- 1 -

LOCAL AND GLOBAL REGISTER PARTITIONING IN A VLIW PROCESSOR

TECHNICAL FIELD

The present invention relates to storage or memory in a processor. More specifically, the present invention relates to a storage having local and global access regions for subinstructions in a Very Long
5 Instruction Word (VLIW) processor.

BACKGROUND ART

One technique for improving the performance of processors is parallel execution of multiple instructions to allow the instruction execution rate to exceed the clock rate. Various types of parallel processors have been developed including Very Long Instruction Word (VLIW) processors that use multiple, independent functional
10 units to execute multiple instructions in parallel. VLIW processors package multiple operations into one very long instruction, the multiple operations being determined by sub-instructions that are applied to the independent functional units. An instruction has a set of fields corresponding to each functional unit. Typical bit lengths of a subinstruction commonly range from 16 to 24 bits per functional unit to produce an instruction length often in a range from 112 to 168 bits.

15 The multiple functional units are kept busy by maintaining a code sequence with sufficient operations to keep instructions scheduled. A VLIW processor often uses a technique called trace scheduling to maintain scheduling efficiency by unrolling loops and scheduling code across basic function blocks. Trace scheduling also improves efficiency by allowing instructions to move across branch points.

Limitations of VLIW processing include limited parallelism, limited hardware resources, and a vast
20 increase in code size. A limited amount of parallelism is available in instruction sequences. Unless loops are unrolled a very large number of times, insufficient operations are available to fill the instructions. Limited hardware resources are a problem, not only because of duplication of functional units but more importantly due to a large increase in memory and register file bandwidth. A large number of read and write ports are necessary for accessing the register file, imposing a bandwidth that is difficult to support without a large cost in the size of
25 the register file and degradation in clock speed. As the number of ports increases, the complexity of the memory system further increases. To allow multiple memory accesses in parallel, the memory is divided into multiple banks having different addresses to reduce the likelihood that multiple operations in a single instruction have conflicting accesses that cause the processor to stall since synchrony must be maintained between the functional units.

30 Code size is a problem for several reasons. The generation of sufficient operations in a nonbranching code fragment requires substantial unrolling of loops, increasing the code size. Also, instructions that are not full may include unused subinstructions that waste code space, increasing code size. Furthermore, the increase in the

- 2 -

size of storages such as the register file increase the number of bits in the instruction for addressing registers in the register file.

A register file with a large number of registers is often used to increase performance of a VLIW processor. A VLIW processor is typically implemented as a deeply pipelined engine with an "in-order" 5 execution model. To attain a high performance a large number of registers is utilized so that the multiple functional units are busy as often as possible.

A large register file has several drawbacks. First, as the number of registers that are directly addressable is increased, the number of bits used to specify the multiple registers within the instruction increases proportionally. For a rich instruction set architecture with, for example, four register specifiers, an additional bit 10 for a register specifier effectively costs four bits per subinstruction (one bit per register specifier). For a VLIW word with four to eight subinstructions, sixteen to thirty-two bits are added for instruction encoding. Second, a register file with many registers occupies a large area. Third, a register file with many registers may create critical timing paths and therefore limit the cycle time of the processor.

What is needed is a technique and processor architecture enhancement that improves the efficiency of 15 instruction coding but still allows access to a large set of architecturally-visible registers.

DISCLOSURE OF INVENTION

A Very Long Instruction Word (VLIW) processor having a plurality of functional units and includes a multi-ported register file that is divided into a plurality of separate register file segments, each of the register file segments being associated to one of the plurality of functional units. The register file segments are partitioned 20 into local registers and global registers. The global registers are read and written by all functional units. The local registers are read and written only by a functional unit associated with a particular register file segment. The local registers and global registers are addressed using register addresses in an address space that is separately defined for a register file segment/ functional unit pair. The global registers are addressed within a selected global register range using the same register addresses for the plurality of register file segment/ 25 functional unit pairs. The local registers in a register file segment are addressed using register addresses in a local register range outside the global register range that are assigned within a single register file segment/ functional unit pair. Register addresses in the local register range are the same for the plurality of register file segment/ functional unit pairs and address registers locally within a register file segment/ functional unit pair.

A VLIW processor utilizes a very long instruction word that includes a plurality of subinstructions. The 30 subinstructions are allocated into positions of the instruction word. The VLIW processor includes a register file that is divided into a plurality of register file segments. The VLIW processor also includes a plurality of functional units, each of which is coupled to and associated with a register file segment of the register file. Each

- 3 -

of the subinstructions executes on a particular functional unit, operating upon operands read from and written to a particular register file segment that is associated with the functional unit. The register file segments include a plurality of registers that are partitioned into global registers and local registers. A global register is a register that is accessed by all of the plurality of functional units and is therefore read and/or written by all functional
5 units. A local register is a register that is only accessible by a particular subinstruction and is only readable and writable by the specific functional unit that executes instructions in the particular subinstruction position.

A register file includes N physical registers. The N-register register file is duplicated into M register file segments, each having a reduced number of read and/or write ports in comparison to a nonduplicated register file, but each having the same number of physical registers. The register file segments are partitioned into N_G
10 global and N_L local register files where N_G plus N_L is equal to N. The register file operates equivalently to having $N_G + (M * N_L)$ total registers available for the M subinstructions. The number of address bits for addressing the $N_G + (M * N_L)$ total registers remains equal to the number of bits B that are used to address $N = 2^B$ registers. The local registers for each of the M register file segments are addressed using the same B-bit values.

In one example, each of M equal to four register file segments includes N equal to 128 registers. The
15 number N_G of global registers is set to 96 while the number N_L of local registers in each register file segment is set to 32. Therefore, the total number of distinct and independent registers is $96 + (4*32) = 224$. The 224 registers are addressed using 7 bits that define an address space from 0-127, rather than the 8 bits that are otherwise required to access 224 registers. In one example, the 96 global registers are addressed using address specifiers 0-95 in all of the four register file segments. Local registers 96-127 in a register file segment 0, local
20 registers 128-159 in register file segment 1, local registers 160-191 in register file segment 2, and local registers 192-223 in register file segment 3 are all addressed using register addresses 96-127.

One address bit is thus saved for each of the four subinstruction positions, a savings of four bits per subinstruction and a savings of 16 bits for a VLIW instruction. The reduction in address bits is highly advantageous in a VLIW processor that includes powerful functional units that execute a large plurality of
25 instructions, each of which is to be encoded in the VLIW instruction word.

In some embodiments, partitioning of the register file is programmable so that the number N_G of global registers and number N_L of local registers is selectable and variable. For example, a register file including four register file segments each having 128 registers may be programmably configured as a flat register file with 128 global registers and 0 local registers with the 128 registers addressed using seven address bits. Alternatively, the
30 four register file segments may be programmably configured, for example, to include 64 global registers and 64 local registers so that the total number of registers is $64 + (4*64) = 320$ registers that are again addressed using 7 bits rather than the 9 bits that would otherwise be required to address 320 registers.

- 4 -

BRIEF DESCRIPTION OF DRAWINGS

The features of the described embodiments are specifically set forth in the appended claims. However, embodiments of the invention relating to both structure and method of operation, may best be understood by referring to the following description and accompanying drawings.

5 **FIGURE 1** is a schematic block diagram illustrating a single integrated circuit chip implementation of a processor in accordance with an embodiment of the present invention.

FIGURE 2 is a schematic block diagram showing the core of the processor.

FIGURE 3 is a schematic block diagram that illustrates an embodiment of the split register file that is suitable for usage in the processor.

10 **FIGURE 4** is a schematic block diagram that shows a logical view of the register file and functional units in the processor.

FIGUREs 5A, 5B, and 5C show a schematic block diagram of a divided or split register file, a high level view of computation elements of a functional unit, and a pictorial view of an instruction format, respectively, which are used to illustrate the difficulty of defining an instruction format with a limited number of
15 instruction bits.

FIGURE 6 is a schematic block diagram showing a register file for a VLIW processor that includes global and local register partitioning.

FIGURE 7 illustrates a schematic block diagram of an SRAM array used for the multi-port split register file.

20 **FIGURE 8A and 8B** are, respectively, a schematic block diagram and a pictorial diagram that illustrate the register file and a memory array insert of the register file.

FIGURE 9 is a schematic block diagram showing an arrangement of the register file into the four register file segments.

FIGURE 10 is a schematic timing diagram that illustrates timing of the processor pipeline.

25 The use of the same reference symbols in different drawings indicates similar or identical items.

MODES FOR CARRYING OUT THE INVENTION

Referring to FIGURE 1, a schematic block diagram illustrates a single integrated circuit chip implementation of a processor 100 that includes a memory interface 102, a geometry decompressor 104, two media processing units 110 and 112, a shared data cache 106, and several interface controllers. The interface controllers support an interactive graphics environment with real-time constraints by integrating fundamental components of memory, graphics, and input/output bridge functionality on a single die. The components are mutually linked and closely linked to the processor core with high bandwidth, low-latency communication channels to manage multiple high-bandwidth data streams efficiently and with a low response time. The interface controllers include a UltraPort Architecture Interconnect (UPA) controller 116 and a peripheral component interconnect (PCI) controller 120. The illustrative memory interface 102 is a direct Rambus dynamic RAM (DRDRAM) controller. The shared data cache 106 is a dual-ported storage that is shared among the media processing units 110 and 112 with one port allocated to each media processing unit. The data cache 106 is four-way set associative, follows a write-back protocol, and supports hits in the fill buffer (not shown). The data cache 106 allows fast data sharing and eliminates the need for a complex, error-prone cache coherency protocol between the media processing units 110 and 112.

The UPA controller 116 maintains a duplicate set of cache tags in the system and performs duplicate tag lookup and main memory initiation in parallel pipelines for each coherent transaction. The approach of the UPA controller 116 attains reduced latency on cache misses and improved utilization of address, datapath, and main memory in comparison to directory-based systems. Directory-based systems maintain coherence states for each data block in main memory and require read-modify-write penalty for every read transaction that reaches main memory. The UPA controller 116 is a centralized system controller that removes the need to place cache coherence logic on the processor 100 and DMA devices, thereby simplifying the circuitry.

The PCI controller 120 is used as the primary system I/O interface for connecting standard, high-volume, low-cost peripheral devices, although other standard interfaces may also be used. The PCI bus effectively transfers data among high bandwidth peripherals and low bandwidth peripherals, such as CD-ROM players, DVD players, and digital cameras.

Two media processing units 110 and 112 are included in a single integrated circuit chip to support an execution environment exploiting thread level parallelism in which two independent threads can execute simultaneously. The threads may arise from any sources such as the same application, different applications, the operating system, or the runtime environment. Parallelism is exploited at the thread level since parallelism is rare beyond four, or even two, instructions per cycle in general purpose code. For example, the illustrative processor 100 is an eight-wide machine with eight execution units for executing instructions. A typical "general-purpose" processing code has an instruction level parallelism of about two so that, on average, most (about six) of the eight execution units would be idle at any time. The illustrative processor 100 employs thread level parallelism and

operates on two independent threads, possibly attaining twice the performance of a processor having the same resources and clock rate but utilizing traditional non-thread parallelism.

Thread level parallelism is particularly useful for JavaTM applications which are bound to have multiple threads of execution. JavaTM methods including "suspend", "resume", "sleep", and the like include effective support for threaded program code. In addition, JavaTM class libraries are thread-safe to promote parallelism. (JavaTM, Sun, Sun Microsystems and the Sun Logo are trademarks or registered trademarks of Sun Microsystems, Inc. in the United States and other countries. All SPARC trademarks, including UltraSPARC I and UltraSPARC II, are used under license and are trademarks of SPARC International, Inc. in the United States and other countries. Products bearing SPARC trademarks are based upon an architecture developed by Sun Microsystems, Inc.) Furthermore, the thread model of the processor 100 supports a dynamic compiler which runs as a separate thread using one media processing unit 110 while the second media processing unit 112 is used by the current application. In the illustrative system, the compiler applies optimizations based on "on-the-fly" profile feedback information while dynamically modifying the executing code to improve execution on each subsequent run. For example, a "garbage collector" may be executed on a first media processing unit 110, copying objects or gathering pointer information, while the application is executing on the other media processing unit 112.

Although the processor 100 shown in FIGURE 1 includes two processing units on an integrated circuit chip, the architecture is highly scaleable so that one to several closely-coupled processors may be formed in a message-based coherent architecture and resident on the same die to process multiple threads of execution. Thus, in the processor 100, a limitation on the number of processors formed on a single die thus arises from capacity constraints of integrated circuit technology rather than from architectural constraints relating to the interactions and interconnections between processors.

Referring to FIGURE 2, a schematic block diagram shows the core of the processor 100. The media processing units 110 and 112 each include an instruction cache 210, an instruction aligner 212, an instruction buffer 214, a pipeline control unit 226, a split register file 216, a plurality of execution units, and a load/store unit 218. In the illustrative processor 100, the media processing units 110 and 112 use a plurality of execution units for executing instructions. The execution units for a media processing unit 110 include three media functional units (MFU) 220 and one general functional unit (GFU) 222. The media functional units 220 are multiple single-instruction-multiple-datatype (MSIMD) media functional units. Each of the media functional units 220 is capable of processing parallel 16-bit components. Various parallel 16-bit operations supply the single-instruction-multiple-datatype capability for the processor 100 including add, multiply-add, shift, compare, and the like. The media functional units 220 operate in combination as tightly-coupled digital signal processors (DSPs). Each media functional unit 220 has an separate and individual sub-instruction stream, but all three media functional units 220 execute synchronously so that the subinstructions progress lock-step through pipeline stages.

- 7 -

The general functional unit 222 is a RISC processor capable of executing arithmetic logic unit (ALU) operations, loads and stores, branches, and various specialized and esoteric functions such as parallel power operations, reciprocal square root operations, and many others. The general functional unit 222 supports less common parallel operations such as the parallel reciprocal square root instruction.

5 The illustrative instruction cache 210 has a 16 Kbyte capacity and includes hardware support to maintain coherence, allowing dynamic optimizations through self-modifying code. Software is used to indicate that the instruction storage is being modified when modifications occur. The 16K capacity is suitable for performing graphic loops, other multimedia tasks or processes, and general-purpose Java™ code. Coherency is maintained by hardware that supports write-through, non-allocating caching. Self-modifying code is supported
10 through explicit use of "store-to-instruction-space" instructions *store2i*. Software uses the *store2i* instruction to maintain coherency with the instruction cache 210 so that the instruction caches 210 do not have to be snooped on every single store operation issued by the media processing unit 110.

15 The pipeline control unit 226 is connected between the instruction buffer 214 and the functional units and schedules the transfer of instructions to the functional units. The pipeline control unit 226 also receives status signals from the functional units and the load/store unit 218 and uses the status signals to perform several control functions. The pipeline control unit 226 maintains a scoreboard, generates stalls and bypass controls.
The pipeline control unit 226 also generates traps and maintains special registers.

20 Each media processing unit 110 and 112 includes a split register file 216, a single logical register file including 128 thirty-two bit registers. The split register file 216 is split into a plurality of register file segments 224 to form a multi-ported structure that is replicated to reduce the integrated circuit die area and to reduce access time. A separate register file segment 224 is allocated to each of the media functional units 220 and the general functional unit 222. In the illustrative embodiment, each register file segment 224 has 128 32-bit registers. The first 96 registers (0-95) in the register file segment 224 are global registers. All functional units can write to the 96 global registers. The global registers are coherent across all functional units (MFU and GFU)
25 so that any write operation to a global register by any functional unit is broadcast to all register file segments 224. Registers 96-127 in the register file segments 224 are local registers. Local registers allocated to a functional unit are not accessible or "visible" to other functional units.

30 The media processing units 110 and 112 are highly structured computation blocks that execute software-scheduled data computation operations with fixed, deterministic and relatively short instruction latencies, operational characteristics yielding simplification in both function and cycle time. The operational characteristics support multiple instruction issue through a pragmatic very large instruction word (VLIW) approach that avoids hardware interlocks to account for software that does not schedule operations properly. Such hardware interlocks are typically complex, error-prone, and create multiple critical paths. A VLIW instruction word always includes one instruction that executes in the general functional unit (GFU) 222 and from zero to three instructions that

- 8 -

execute in the media functional units (MFU) 220. A MFU instruction field within the VLIW instruction word includes an operation code (opcode) field, three source register (or immediate) fields, and one destination register field.

Instructions are executed in-order in the processor 100 but loads can finish out-of-order with respect to other instructions and with respect to other loads, allowing loads to be moved up in the instruction stream so that data can be streamed from main memory. The execution model eliminates the usage and overhead resources of an instruction window, reservation stations, a re-order buffer, or other blocks for handling instruction ordering. Elimination of the instruction ordering structures and overhead resources is highly advantageous since the eliminated blocks typically consume a large portion of an integrated circuit die. For example, the eliminated blocks consume about 30% of the die area of a Pentium II processor.

To avoid software scheduling errors, the media processing units 110 and 112 are high-performance but simplified with respect to both compilation and execution. The media processing units 110 and 112 are most generally classified as a simple 2-scalar execution engine with full bypassing and hardware interlocks on load operations. The instructions include loads, stores, arithmetic and logic (ALU) instructions, and branch instructions so that scheduling for the processor 100 is essentially equivalent to scheduling for a simple 2-scalar execution engine for each of the two media processing units 110 and 112.

The processor 100 supports full bypasses between the first two execution units within the media processing unit 110 and 112 and has a scoreboard in the general functional unit 222 for load operations so that the compiler does not need to handle nondeterministic latencies due to cache misses. The processor 100 scoreboards long latency operations that are executed in the general functional unit 222, for example a reciprocal square-root operation, to simplify scheduling across execution units. The scoreboard (not shown) operates by tracking a record of an instruction packet or group from the time the instruction enters a functional unit until the instruction is finished and the result becomes available. A VLIW instruction packet contains one GFU instruction and from zero to three MFU instructions. The source and destination registers of all instructions in an incoming VLIW instruction packet are checked against the scoreboard. Any true dependencies or output dependencies stall the entire packet until the result is ready. Use of a scoreboxed result as an operand causes instruction issue to stall for a sufficient number of cycles to allow the result to become available. If the referencing instruction that provokes the stall executes on the general functional unit 222 or the first media functional unit 220, then the stall only endures until the result is available for intra-unit bypass. For the case of a *load* instruction that hits in the data cache 106, the stall may last only one cycle. If the referencing instruction is on the second or third media functional units 220, then the stall endures until the result reaches the writeback stage in the pipeline where the result is bypassed in transmission to the split register file 216.

- 9 -

The scoreboard automatically manages load delays that occur during a load hit. In an illustrative embodiment, all loads enter the scoreboard to simplify software scheduling and eliminate NOPs in the instruction stream.

The scoreboard is used to manage most interlocks between the general functional unit 222 and the media functional units 220. All loads and non-pipelined long-latency operations of the general functional unit 222 are scoreboxed. The long-latency operations include division *idiv*, *fdiv* instructions, reciprocal square root *frcsqrt*, *precsgrt* instructions, and power *ppower* instructions. None of the results of the media functional units 220 is scoreboxed. Non-scoreboxed results are available to subsequent operations on the functional unit that produces the results following the latency of the instruction.

The illustrative processor 100 has a rendering rate of over fifty million triangles per second without accounting for operating system overhead. Therefore, data feeding specifications of the processor 100 are far beyond the capabilities of cost-effective memory systems. Sufficient data bandwidth is achieved by rendering of compressed geometry using the geometry decompressor 104, an on-chip real-time geometry decompression engine. Data geometry is stored in main memory in a compressed format. At render time, the data geometry is fetched and decompressed in real-time on the integrated circuit of the processor 100. The geometry decompressor 104 advantageously saves memory space and memory transfer bandwidth. The compressed geometry uses an optimized generalized mesh structure that explicitly calls out most shared vertices between triangles, allowing the processor 100 to transform and light most vertices only once. In a typical compressed mesh, the triangle throughput of the transform-and-light stage is increased by a factor of four or more over the throughput for isolated triangles. For example, during processing of triangles, multiple vertices are operated upon in parallel so that the utilization rate of resources is high, achieving effective spatial software pipelining. Thus operations are overlapped in time by operating on several vertices simultaneously, rather than overlapping several loop iterations in time. For other types of applications with high instruction level parallelism, high trip count loops are software-pipelined so that most media functional units 220 are fully utilized.

Referring to FIGURE 3, a schematic block diagram illustrates an embodiment of the split register file 216 that is suitable for usage in the processor 100. The split register file 216 supplies all operands of processor instructions that execute in the media functional units 220 and the general functional units 222 and receives results of the instruction execution from the execution units. The split register file 216 operates as an interface to the geometry decompressor 104. The split register file 216 is the source and destination of store and load operations, respectively.

In the illustrative processor 100, the split register file 216 in each of the media processing units 110 and 112 has 128 registers. Graphics processing places a heavy burden on register usage. Therefore, a large number of registers is supplied by the split register file 216 so that performance is not limited by loads and stores or handling of intermediate results including graphics "fills" and "spills". The illustrative split register file 216

- 10 -

includes twelve read ports and five write ports, supplying total data read and write capacity between the central registers of the split register file 216 and all media functional units 220, the general functional unit 222 and the load/store unit 218 that is connected to the general functional unit 222. The five write ports include one 64-bit write port that is dedicated to load operations. The remaining four write ports are 32 bits wide and are used to 5 write operations of the general functional unit 222 and the media functional units 220.

Total read and write capacity promotes flexibility and facility in programming both of hand-coded routines and compiler-generated code.

Large, multiple-ported register files are typically metal-limited so that the register area is proportional with the square of the number of ports. A sixteen port file is roughly proportional in size and speed to a value of 10 256. The illustrative split register file 216 is divided into four register file segments 310, 312, 314, and 316, each having three read ports and four write ports so that each register file segment has a size and speed proportional to 49 for a total area for the four segments that is proportional to 196. The total area is therefore potentially smaller and faster than a single central register file. Write operations are fully broadcast so that all files are maintained coherent. Logically, the split register file 216 is no different from a single central register file. However, from 15 the perspective of layout efficiency, the split register file 216 is highly advantageous, allowing for reduced size and improved performance through faster access.

The new media data that is operated upon by the processor 100 is typically heavily compressed. Data transfers are communicated in a compressed format from main memory and input/output devices to pins of the processor 100, subsequently decompressed on the integrated circuit holding the processor 100, and passed to the 20 split register file 216.

Splitting the register file into multiple segments in the split register file 216 in combination with the character of data accesses in which multiple bytes are transferred to the plurality of execution units concurrently, results in a high utilization rate of the data supplied to the integrated circuit chip and effectively leads to a much higher data bandwidth than is supported on general-purpose processors. The highest data bandwidth requirement 25 is therefore not between the input/output pins and the central processing units, but is rather between the decompressed data source and the remainder of the processor. For graphics processing, the highest data bandwidth requirement is between the geometry decompressor 104 and the split register file 216. For video decompression, the highest data bandwidth requirement is internal to the split register file 216. Data transfers between the geometry decompressor 104 and the split register file 216 and data transfers between various 30 registers of the split register file 216 can be wide and run at processor speed, advantageously delivering a large bandwidth. In addition, the split register file 216 can be multiported which further increases total bandwidth.

The register file 216 is a focal point for attaining the very large bandwidth of the processor 100. The processor 100 transfers data using a plurality of data transfer techniques. In one example of a data transfer

- 11 -

technique, cacheable data is loaded into the split register file 216 through normal load operations at a low rate of up to eight bytes per cycle. In another example, streaming data is transferred to the split register file 216 through group load operations which transfer thirty-two bytes from memory directly into eight consecutive 32-bit registers. For example, the processor 100 utilizes the streaming data operation to receive compressed video data 5 for decompression.

Compressed graphics data is received via a direct memory access (DMA) unit in the geometry decompressor 104. The compressed graphics data is decompressed by the geometry decompressor 104 and loaded at a high bandwidth rate into the split register file 216 via group load operations that are mapped to the geometry decompressor 104.

10 Load operations are non-blocking and scoreboarded so that a long latency inherent to loads can be hidden by early scheduling.

General purpose applications often fail to exploit the large register file 216. Statistical analysis shows that compilers do not effectively use the large number of registers in the split register file 216. However, aggressive in-lining techniques that have traditionally been restricted due to the limited number of registers in 15 conventional systems may be advantageously used in the processor 100 to exploit the large number of registers in the split register file 216. In a software system that exploits the large number of registers in the processor 100, the complete set of registers is saved upon the event of a thread (context) switch. When only a few registers of the entire set of registers is used, saving all registers in the full thread switch is wasteful. Waste is avoided in the processor 100 by supporting individual marking of registers. Octants of the thirty-two registers can be marked as 20 "dirty" if used, and are consequently saved conditionally.

In various embodiments, the split register file 216 is leveraged by dedicating fields for globals, trap registers, and the like.

Referring to FIGURE 4, a schematic block diagram shows a logical view of the register file 216 and functional units in the processor 100. The physical implementation of the core processor 100 is simplified by 25 replicating a single functional unit to form the three media functional units 220. The media functional units 220 include circuits that execute various arithmetic and logical operations including general-purpose code, graphics code, and video-image-speech (VIS) processing. VIS processing includes video processing, image processing, digital signal processing (DSP) loops, speech processing, and voice recognition algorithms, for example.

Referring to FIGUREs 5A, 5B, and 5C, a schematic block diagram of a divided or split register file, a 30 high level view of computation elements of a functional unit, and a pictorial view of an instruction format, respectively are used to illustrate the difficulty of defining an instruction format with a limited number of instruction bits. FIGURE 5A shows a schematic block diagram of a decoder 502 that decodes four

- 12 -

subinstructions of a very long instruction word. Each of the four decoders applies control signals to one of four register file segments 510, 512, 514, and 516. Each of the register file segments is coupled to and associated with a functional unit. In the illustrative embodiment, a first register file segment 510 is coupled to and associated with a general functional unit 520. Second, third, and fourth register file segments 512, 514, and 516 are respectively coupled to and associated with media functional units 522, 524, and 526.

FIGURE 5B shows an example of a VLIW subinstruction, specifically a multiply-add (muladd) instruction and relates execution of the muladd instruction to computation blocks in a functional unit. The muladd instruction specifies four register specifiers designating data that is operated upon by the functional unit. The muladd instruction specifies three source operands R_A , R_B , and R_C , and one destination operand R_D . The functional unit includes a multiplier 530 that multiplies the source operands R_A and R_B to generate a product. The functional unit also includes an adder 532 that receives the product from the multiplier 530 and adds the product and the source operand R_C to produce a sum that is transferred to the destination register operand R_D .

For a register file in which the register file segments include $N = 2^M$ registers, for example, M bits are used to uniquely specify a particular register so that $4 \cdot M$ are needed to uniquely specify the four registers addressed in a single subinstruction.

FIGURE 5C depicts a subinstruction storage for instructions such as the muladd instruction. Resource size and speed constraints are imposed on instruction storage so that the number of bits in a subinstruction are limited. The four register specifiers for the subinstruction use nearly the entire capacity of the subinstruction storage. For example, a register file segment that includes 128 bits has registers that are uniquely addressed using seven address bits. Addressing of four registers consumes $7 \cdot 4 = 28$ bits. For a subinstruction size constrained to 32 bits, only four bits remain for specifying an operation code or other operational information for controlling execution.

The illustrative VLIW processor partitions the register file into local and global registers to conserve address bits in a very long instruction word to reduce the size of the register file and accelerate access time.

Referring to FIGURE 6, a schematic block diagram shows a register file 600 for a VLIW processor 100 that includes global and local register partitioning. The Very Long Instruction Word (VLIW) processor has a plurality of functional units including three media functional units 622, 624, and 626, and a general functional unit 620. The processor 100 also includes a multi-ported register file 600 that is divided into a plurality of separate register file segments 610, 612, 614, and 616, each of the register file segments being associated to one of the plurality of functional units. The register file segments 610, 612, 614, and 616 are partitioned into local registers and global registers. The global registers are read and written by all functional units 620, 622, 624, and 626. The local registers are read and written only by a functional unit associated with a particular register file segment. The local registers and global registers are addressed using register addresses in an address space that is

- 13 -

separately defined for a register file segment/ functional unit pair including register file segment 610/ general functional unit 620, register file segment 612/ media functional unit 622, register file segment 614/ media functional unit 624, and register file segment 616/ media functional unit 626.

The global registers are addressed within a selected global register range using the same register

- 5 addresses for the plurality of register file segment/ functional unit pairs, for example, global registers 0-95. The local registers in a register file segment are addressed using register addresses in a local register range outside the global register range, for example addresses 96-127, that are assigned within a single register file segment/ functional unit pair. Register addresses 96-127 applied to the register file segments in the local register range are the same for the plurality of register file segment/ functional unit pairs and address registers locally within a 10 register file segment/ functional unit pair. The register specifiers of the local registers, as defined external to processor, do not overlap but instead have distinct and different specifiers. For example, in one embodiment, external to the processor 100, the 96 global registers are addressed using address specifiers 0-95 in all of the four register file segments. Local registers 96-127 in the register file segment 610, local registers 128-159 in register file segment 612, local registers 160-191 in register file segment 614, and local registers 192-223 in register file 15 segment 616 are all addressed using register addresses 96-127. In this example, the total number of distinct and independent registers is $96 + (4*32) = 224$. The 224 registers are addressed using 7 bits that define an address space from 0-127, rather than the 8 bits that are otherwise required to access 224 registers.

Global and local register partitioning advantageously leverages the information content of register specifier bits in an instruction word by inherently communicating information by position dependence within a 20 VLIW instruction group. The positioning of a register specifier in the instruction word thus communicates addressing information. The additional information allows a compiler or programmer to specify more registers in fewer bits than have been specified conventionally.

One address bit is thus saved for each of the four subinstruction positions, a savings of four bits per subinstruction and a savings of 16 bits per VLIW instruction. The reduction in address bits is highly 25 advantageous in a VLIW processor that includes powerful functional units that execute a large plurality of instructions, each of which is to be encoded in the VLIW instruction word.

In general embodiments, the register file 600 includes N physical registers. The N-register register file 600 is duplicated into M register file segments 610, 612, 614, and 616, each having a reduced number of read and/or write ports in comparison to a nonduplicated register file, but each having the same number of physical 30 registers. The register file segments are partitioned into N_G global and N_L local register files where N_G plus N_L is equal to N. The register file operates equivalently to having $N_G + (M * N_L)$ total registers available for the M functional units. The number of address bits for addressing the $N_G + (M * N_L)$ total registers remains equal to the number of bits B that are used to address $N = 2^B$ registers. The local registers for each of the M register file segments are addressed using the same B-bit values.

In some embodiments, partitioning of the register file 600 is programmable so that the number N_G of global registers and number N_L of local registers is selectable and variable. For example, a register file including four register file segments each having 128 registers may be programmably configured as a flat register file with 128 global registers and 0 local registers with the 128 registers addressed using seven address bits. Alternatively, 5 the four register file segments may be programmably configured, for example, to include 64 global registers and 64 local registers so that the total number of registers is $64 + (4 \times 64) = 320$ registers that are again addressed using 7 bits rather than the 9 bits that would otherwise be required to address 320 registers.

Referring to FIGURE 7, a schematic block diagram depicts an embodiment of the multiport register file 216. A plurality of read address buses RA1 through RAN carry read addresses that are applied to decoder ports 816-1 through 816-N, respectively. Decoder circuits are well known to those of ordinary skill in the art, and any of several implementations could be used as the decoder ports 816-1 through 816-N. When an address is presented to any of decoder ports 816-1 through 816-N, the address is decoded and a read address signal is transmitted by a decoder port 816 to a register in a memory cell array 818. Data from the memory cell array 818 is output using output data drivers 822. Data is transferred to and from the memory cell array 818 under control 15 of control signals carried on some of the lines of the buses of the plurality of read address buses RA1 through RAN.

Referring to FIGURE 8A and 8B, a schematic block diagram and a pictorial diagram, respectively, illustrate the register file 216 and a memory array insert 910. The register file 216 is connected to a four functional units 920, 922, 924, and 926 that supply information for performing operations such as arithmetic, 20 logical, graphics, data handling operations and the like. The illustrative register file 216 has twelve read ports 930 and four write ports 932. The twelve read ports 930 are illustratively allocated with three ports connected to each of the four functional units. The four write ports 932 are connected to receive data from all of the four functional units.

The register file 216 includes a decoder, as is shown in FIGURE 6, for each of the sixteen read and 25 write ports. The register file 216 includes a memory array 940 that is partially shown in the insert 710 illustrated in FIGURE 8B and includes a plurality of word lines 944 and bit lines 946. The word lines 944 and bit lines 946 are simply a set of wires that connect transistors (not shown) within the memory array 940. The word lines 944 select registers so that a particular word line selects a register of the register file 216. The bit lines 946 are a second set of wires that connect the transistors in the memory array 940. Typically, the word lines 944 and bit 30 lines 946 are laid out at right angles. In the illustrative embodiment, the word lines 944 and the bit lines 946 are constructed of metal laid out in different planes such as a metal 2 layer for the word lines 944 and a metal 3 layer for the bit lines 946. In other embodiments, bit lines and word lines may be constructed of other materials, such as polysilicon, or can reside at different levels than are described in the illustrative embodiment, that are known in the art of semiconductor manufacture. In the illustrative example, the word lines 944 are separated by a

- 15 -

distance of about $1\mu\text{m}$ and the bit lines 946 are separated by approximately $1\mu\text{m}$. Other circuit dimensions may be constructed for various processes. The illustrative example shows one bit line per port, other embodiments may use multiple bit lines per port.

When a particular functional unit reads a particular register in the register file 216, the functional unit
5 sends an address signal via the read ports 930 that activates the appropriate word lines to access the register. In a register file having a conventional structure and twelve read ports, each cell, each storing a single bit of information, is connected to twelve word lines to select an address and twelve bit lines to carry data read from the address.

The four write ports 932 address registers in the register file using four word lines 944 and four bit lines
10 946 connected to each cell. The four word lines 944 address a cell and the four bit lines 946 carry data to the cell.

Thus, if the illustrative register file 216 were laid out in a conventional manner with twelve read ports 930 and four write ports 932 for a total of sixteen ports and the ports were $1\mu\text{m}$ apart, one memory cell would have an integrated circuit area of $256\mu\text{m}^2$ (16×16). The area is proportional to the square of the number of ports.

15 The register file 216 is alternatively implemented to perform single-ended reads and/or single-ended writes utilizing a single bit line per port per cell, or implemented to perform differential reads and/or differential writes using two bit lines per port per cell.

However, in this embodiment the register file 216 is not laid out in the conventional manner and instead is split into a plurality of separate and individual register file segments 224. Referring to FIGURE 9, a
20 schematic block diagram shows an arrangement of the register file 216 into the four register file segments 224. The register file 216 remains operational as a single logical register file in the sense that the four of the register file segments 224 contain the same number of registers and the same register values as a conventional register file of the same capacity that is not split. The separated register file segments 224 differ from a register file that is not split through elimination of lines that would otherwise connect ports to the memory cells. Accordingly,
25 each register file segment 224 has connections to only three of the twelve read ports 930, lines connecting a register file segment to the other nine read ports are eliminated. All writes are broadcast so that each of the four register file segments 224 has connections to all four write ports 932. Thus each of the four register file segments 224 has three read ports and four write ports for a total of seven ports. The individual cells are connected to seven word lines and seven bit lines so that a memory array with a spacing of $1\mu\text{m}$ between lines has an area of
30 approximately $49\mu\text{m}^2$. In the illustrative embodiment, the four register file segments 224 have an area proportional to seven squared. The total area of the four register file segments 224 is therefore proportional to 49 times 4, a total of 196.

- 16 -

The split register file thus advantageously reduces the area of the memory array by a ratio of approximately 256/196 (1.3X or 30%). The reduction in area further advantageously corresponds to an improvement in speed performance due to a reduction in the length of the word lines 944 and the bit lines 946 connecting the array cells that reduces the time for a signal to pass on the lines. The improvement in speed
5 performance is highly advantageous due to strict time budgets that are imposed by the specification of high-performance processors and also to attain a large capacity register file that is operational at high speed. For example, the operation of reading the register file 216 typically takes place in a single clock cycle. For a processor that executes at 500 MHz, a cycle time of two nanoseconds is imposed for accessing the register file 216. Conventional register files typically only have up to about 32 registers in comparison to the 128 registers in
10 the illustrative register file 216 of the processor 100. A register file 216 that is substantially larger than the register file in conventional processors is highly advantageous in high-performance operations such as video and graphic processing. The reduced size of the register file 216 is highly useful for complying with time budgets in a large capacity register file.

In some embodiments, the area of a register file is further reduced by using a special memory cell for
15 the local registers that have write port connections to the functional unit that is locally associated with the register file segment. Thus the local registers are only written by the local functional unit. In the illustrative register file 216, the special cell for local registers has only a single write port, reducing the number of word lines to four. The number of bit lines connected to the local registers is also reduced to four, allowing further compactness to the cell.

20 Referring to FIGURE 10, a simplified schematic timing diagram illustrates timing of the processor pipeline 1100. The pipeline 1100 includes nine stages including three initiating stages, a plurality of execution phases, and two terminating stages. The three initiating stages are optimized to include only those operations necessary for decoding instructions so that jump and call instructions, which are pervasive in the JavaTM language, execute quickly. Optimization of the initiating stages advantageously facilitates branch prediction
25 since branches, jumps, and calls execute quickly and do not introduce many bubbles.

The first of the initiating stages is a fetch stage 1110 during which the processor 100 fetches instructions from the 16Kbyte two-way set-associative instruction cache 210. The fetched instructions are aligned in the instruction aligner 212 and forwarded to the instruction buffer 214 in an align stage 1112, a second stage of the initiating stages. The aligning operation properly positions the instructions for storage in a particular segment of
30 the four register file segments 310, 312, 314, and 316 and for execution in an associated functional unit of the three media functional units 220 and one general functional unit 222. In a third stage, a decoding stage 1114 of the initiating stages, the fetched and aligned VLIW instruction packet is decoded and the scoreboard (not shown) is read and updated in parallel. The four register file segments 310, 312, 314, and 316 each holds either floating-point data or integer data. The register files are read in the decoding (D) stage.

- 17 -

Following the decoding stage 1114, the execution stages are performed. The two terminating stages include a trap-handling stage 1160 and a write-back stage 1162 during which result data is written-back to the split register file 216.

While the invention has been described with reference to various embodiments, it will be understood that these embodiments are illustrative and that the scope of the invention is not limited to them. Many variations, modifications, additions and improvements of the embodiments described are possible. For example, those skilled in the art will readily implement the steps necessary to provide the structures and methods disclosed herein, and will understand that the process parameters, materials, and dimensions are given by way of example only and can be varied to achieve the desired structure as well as modifications which are within the scope of the invention. Variations and modifications of the embodiments disclosed herein may be made based on the description set forth herein, without departing from the scope and spirit of the invention as set forth in the following claims. For example, the described structure and operating method may be applied to a register file and processor of any suitable size. The register file may be much larger or much smaller than the described 128 registers. The very long instruction word may include any suitable number of subinstructions.

Similarly, although the illustrative register file has one bit line per port, in other embodiments more bit lines may be allocated for a port. The described word lines and bit lines are formed of a metal. In other examples, other conductive materials such as doped polysilicon may be employed for interconnects. The described register file uses single-ended reads and writes so that a single bit line is employed per bit and per port. In other processors, differential reads and writes with dual-ended sense amplifiers may be used so that two bit lines are allocated per bit and per port, resulting in a bigger pitch. Dual-ended sense amplifiers improve memory fidelity but greatly increase the size of a memory array, imposing a heavy burden on speed performance. Thus the advantages attained by the described register file structure are magnified for a memory using differential reads and writes. The spacing between bit lines and word lines is described to be approximately 1 μ m. In some processors, the spacing may be greater than 1 μ m. In other processors the spacing between lines is less than 1 μ m.

WE CLAIM

- 1 1. A processor comprising:
2 a plurality of functional units; and
3 a register file characterized in that the register file is divided into a plurality of register file segments,
4 ones of the plurality of register file segments being coupled to and associated with ones of the
5 plurality of functional units, the register file segments being partitioned into global registers
6 and local registers, the global registers that are accessible by the plurality of functional units,
7 the local registers being accessible by the functional unit associated with the register file
8 segment containing the local registers.
- 1 2. A processor comprising:
2 a decoder for decoding a very long instruction word including a plurality of subinstructions, the
3 subinstructions being allocated into positions of the instruction word;
4 a register file coupled to the decoder and divided into a plurality of register file segments; and
5 a plurality of functional units, ones of the plurality of functional units being coupled to an associated
6 with respective ones of the register file segments, ones of the plurality of subinstructions being
7 executable upon respective ones of the plurality of functional units, operating upon operands
8 accessible to the register file segment associated with the functional unit of the plurality of
9 functional units, the register file segments including a plurality of registers that are partitioned
10 into global registers and local registers, the global registers being accessible by the plurality of
11 functional units, the local registers in one of the register file segments being accessible by the
12 functional unit associated with the register file segment.
- 1 3. A processor according to either Claim 1 or Claim 2 wherein:
2 the processor is a Very Long Instruction Word (VLIW) processor.
- 1 4. A processor according to Claim 1 or Claim 2 wherein:
2 the local registers and global registers are addressed using register addresses in an address space that is
3 defined for a register file segment/ functional unit pair.
- 1 5. A processor according to Claim 1 or Claim 2 wherein:
2 the register file is a multi-ported register file.

- 1 6. A processor according to Claim 1 or Claim 2 wherein:
2 the local registers in a register file segment are addressed using register addresses in a local register
3 range outside the global register range that are assigned within a single register file segment/
4 functional unit pair.

- 1 7. A processor according to Claim 1 or Claim 2 wherein:
2 register addresses in the local register range are the same for the plurality of register file segment/
3 functional unit pairs and address registers locally within a register file segment/ functional unit
4 pair.

- 1 8. A processor according to Claim 1 or Claim 2 wherein:
2 the register file includes N physical registers and is duplicated into M register file segments, the register
3 file segments having a reduced number of read and/or write ports in comparison to a
4 nonduplicated register file, but each having the same number of physical registers.

- 1 9. A processor according to Claim 8 wherein:
2 the register file segments are partitioned into N_G global and N_L local register files where N_G plus N_L is
3 equal to N, the register file operating equivalently to a register file having $N_G + (M * N_L)$ total
4 registers available for the M functional units, the number of address bits for addressing the N_G
5 + $(M * N_L)$ total registers being equal to the number of bits B that are used to address $N = 2^B$
6 registers, the local registers for ones of the M register file segments are addressed using the
7 same B-bit values.

- 1 10. A processor according to Claim 9 wherein:
2 partitioning of the register file is programmable so that the number N_G of global registers and number
3 N_L of local registers is selectable and variable.

- 1 11. A processor according to Claim 1 or Claim 2 wherein the register file is a storage array structure
2 having R read ports and W write ports comprising:
3 a plurality of storage array storages;
4 the storage array storages having a reduced number of read ports so that the total number of read ports
5 for the plurality of storage array storages is R read ports; and
6 the storage array storages having W write ports.

1 12. A processor according to Claim 11 wherein:
2 the storage array structure is a sixteen port structure with twelve read ports and five write ports; and
3 the plurality of storage array storages includes four storage array storages each having three read ports
4 and five write ports.

1 13. A processor according to Claim 11 wherein:
2 the storage array structure is a sixteen port structure with twelve read ports and four write ports; and
3 the plurality of storage array storages includes four storage array storages each having three read ports
4 and four write ports.

1 14. A processor according to Claim 11 wherein:
2 the writes are fully broadcast so that all of the storage array storages are held coherent.

1 15. A processor according to Claim 11 wherein:
2 storage array storages include storage cells having a plurality of word lines and a plurality of bit lines,
3 the word lines being formed in one metal interconnect layer, the bits lines being formed in a
4 second metal interconnect layer.

1 16. A method of operating a processor comprising:
2 operating a plurality of functional units; and
3 dividing a register file into a plurality of register file segments;
4 coupling and associating ones of the plurality of register file segments with ones of the plurality of
5 functional units;
6 partitioning the register file segments into global registers and local registers;
7 accessing the global registers by the plurality of functional units;
8 accessing the local registers by the functional unit associated with the register file segment containing
9 the local registers.

1 17. A method according to Claim 16 further comprising:
2 addressing the local registers and global registers using register addresses in an address space that is
3 defined for a register file segment/ functional unit pair.

1 18. A method according to Claim 16 further comprising:
2 addressing the local registers in a register file segment using register addresses in a local register range
3 outside the global register range that are assigned within a single register file segment/
4 functional unit pair.

1 19. A method according to Claim 16 further comprising:
2 addressing the local register range the same for the plurality of register file segment/ functional unit
3 pairs and address registers locally within a register file segment/ functional unit pair.

1 20. A method according to Claim 16 further comprising:
2 including N physical registers in the register file;
3 duplicating the physical registers into M register file segments, the register file segments having a
4 reduced number of read and/or write ports in comparison to a nonduplicated register file, but
5 each having the same number of physical registers.

1 21. A method according to Claim 20 further comprising:
2 partitioning the register file segments into N_G global and N_L local register files where N_G plus N_L is
3 equal to N;
4 operating the register file equivalently to a register file having N_G + (M * N_L) total registers available for
5 the M functional units, the number of address bits for addressing the N_G + (M * N_L) total
6 registers being equal to the number of bits B that are used to address N = 2^B registers; and
7 addressing the local registers for ones of the M register file segments using the same B-bit values.

1 22. A method according to Claim 20 further comprising:
2 programmably partitioning the register file so that the number N_G of global registers and number N_L of
3 local registers is selectable and variable.

1/7

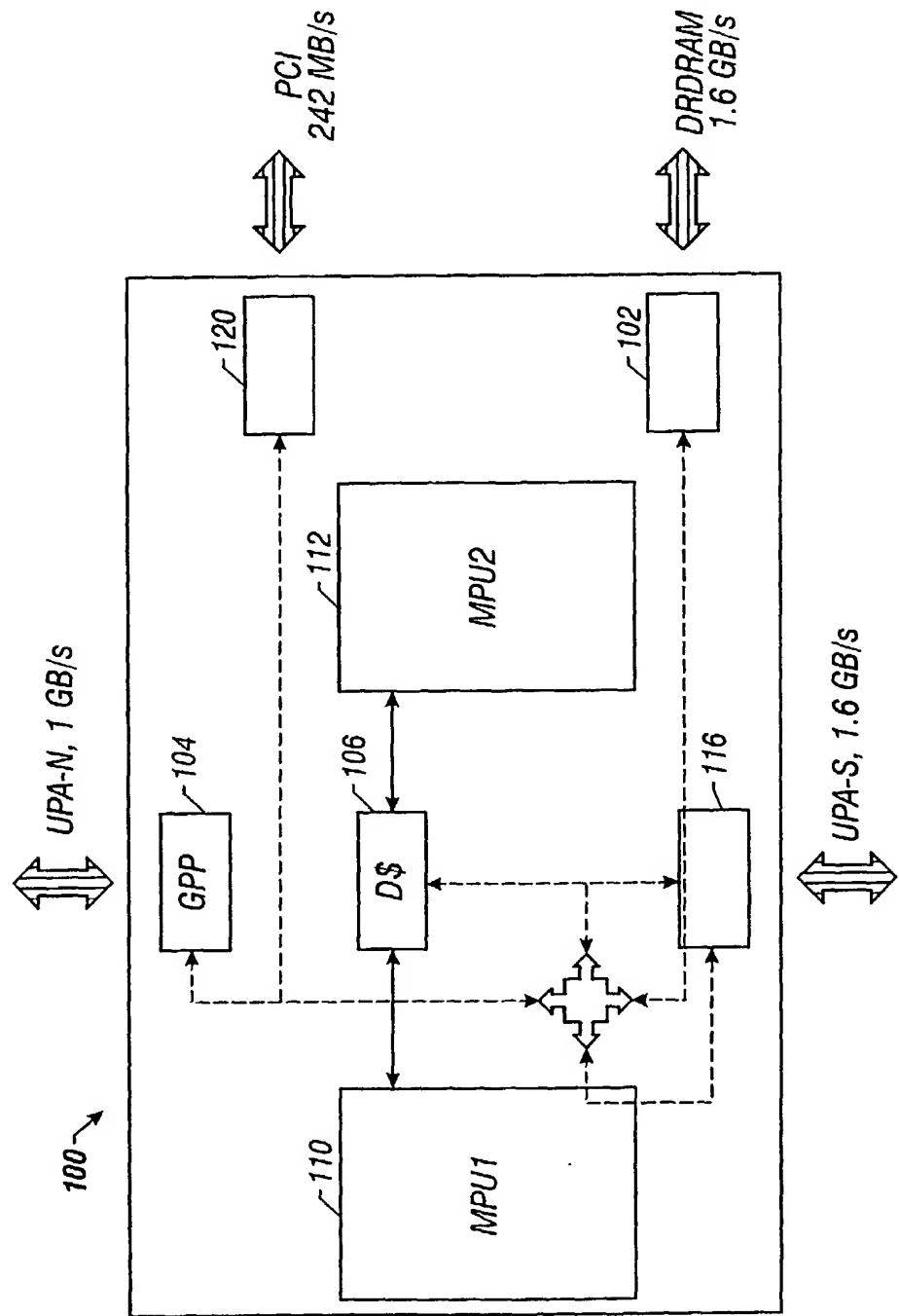


FIG. 1

2/7

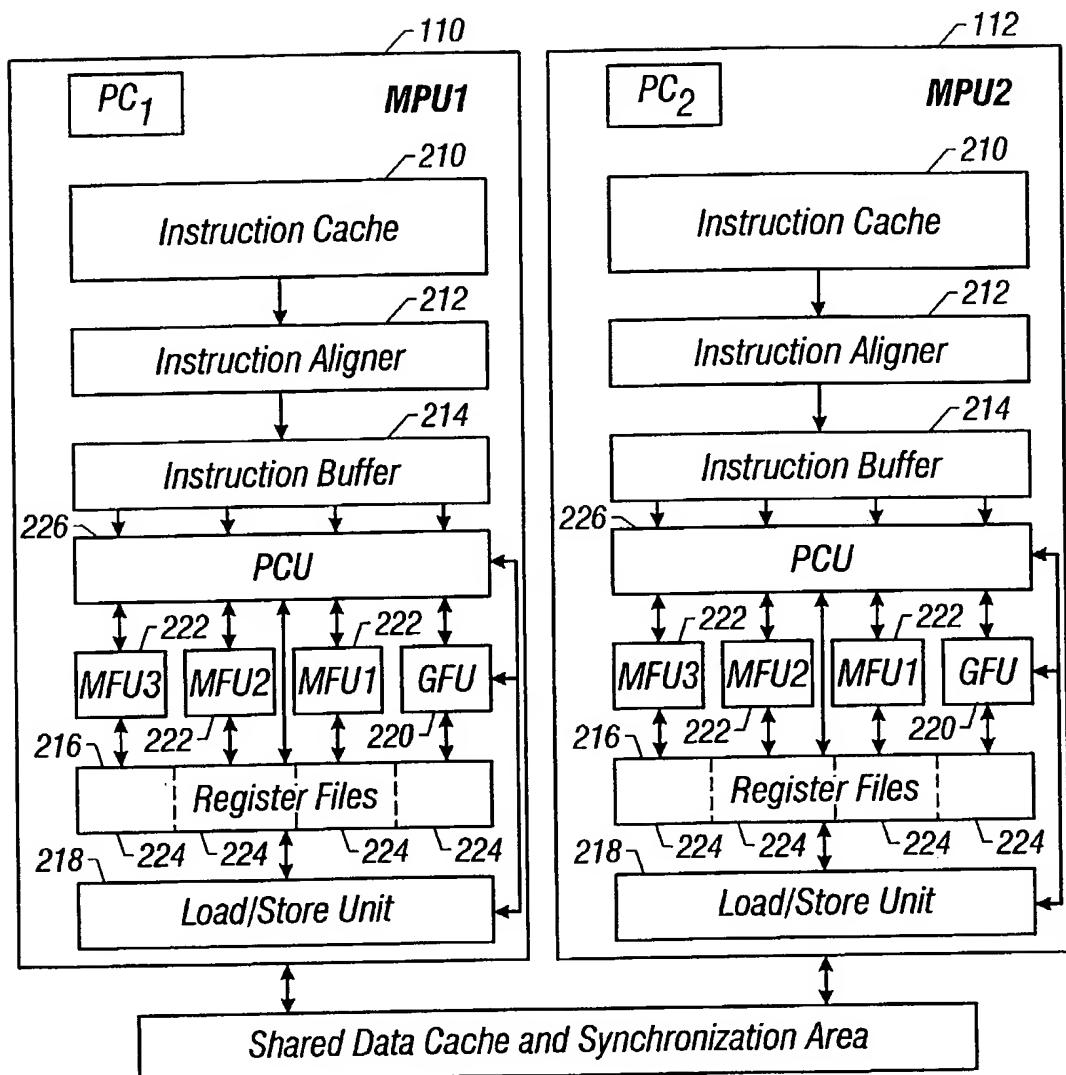


FIG. 2

3/7

216

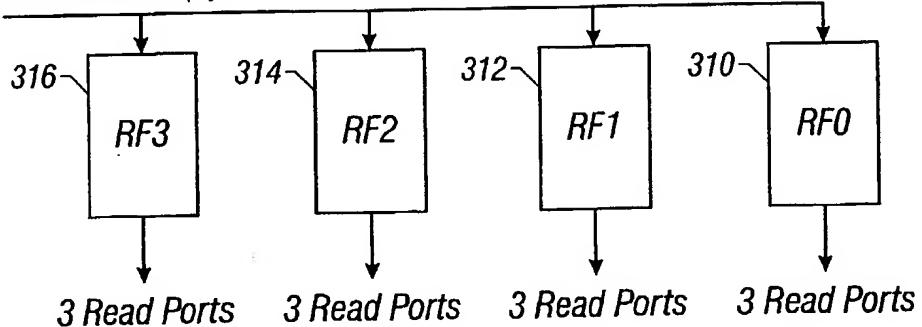
Broadcast Writes (5)

FIG. 3

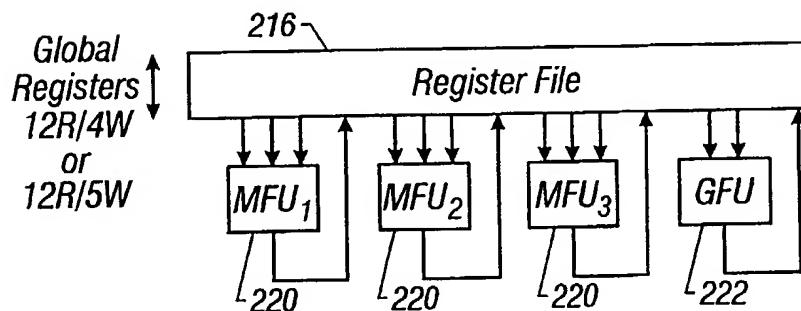


FIG. 4

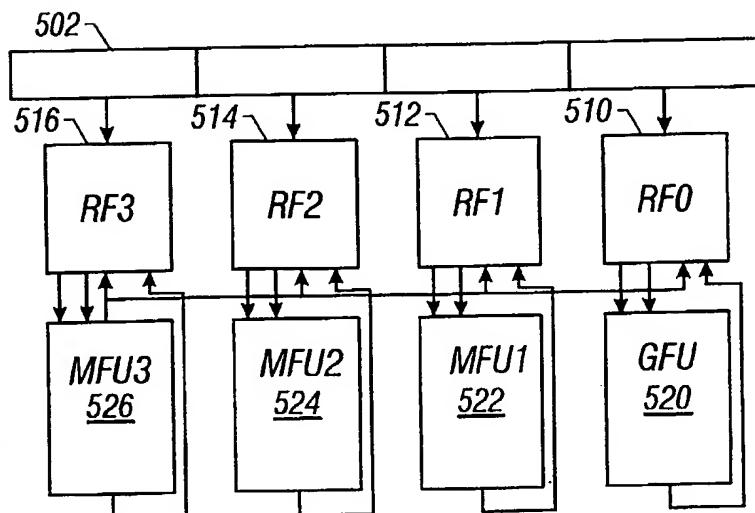
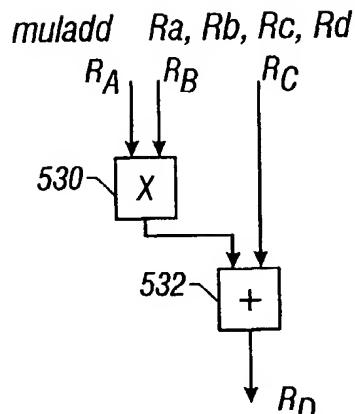
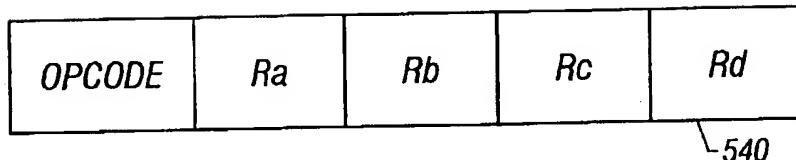
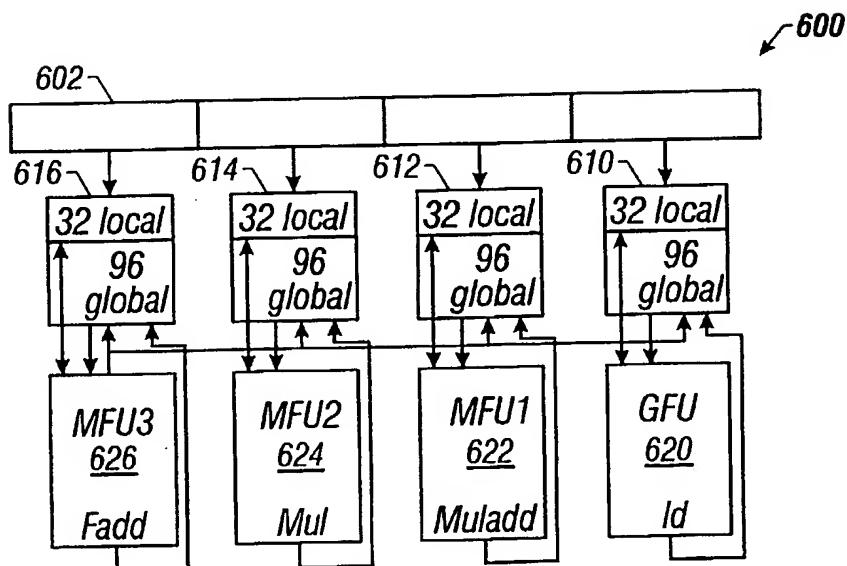


FIG. 5A

4/7

**FIG. 5B****FIG. 5C****FIG. 6**

5/7

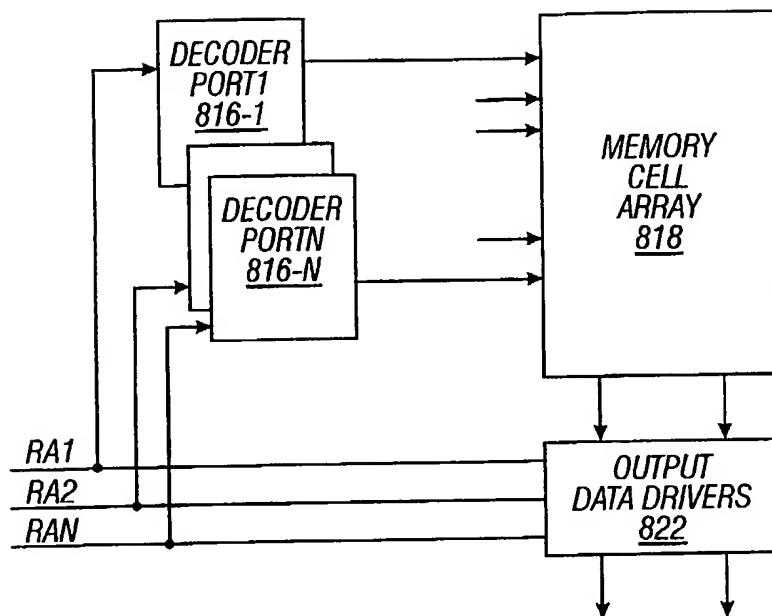


FIG. 7

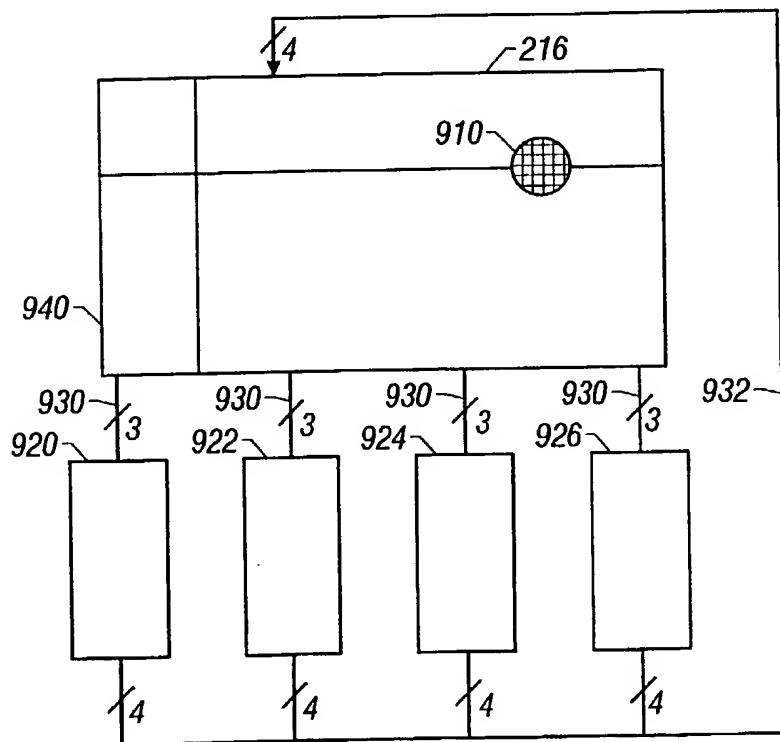


FIG. 8A

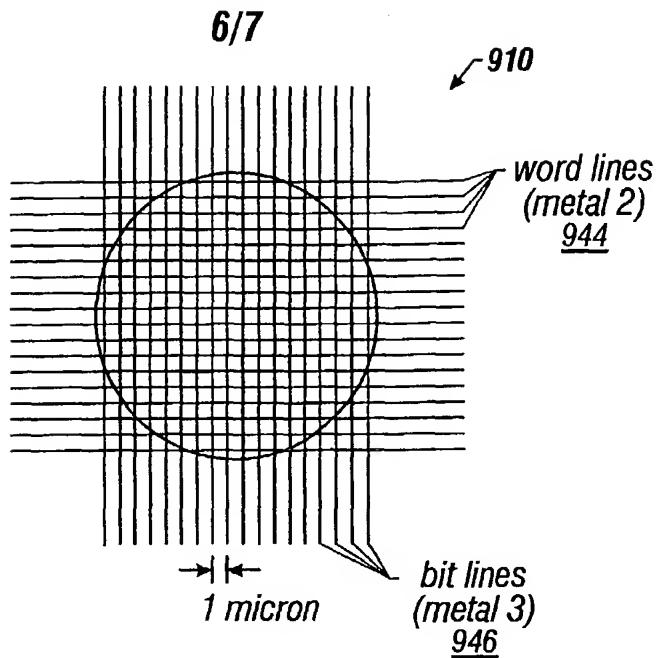


FIG. 8B

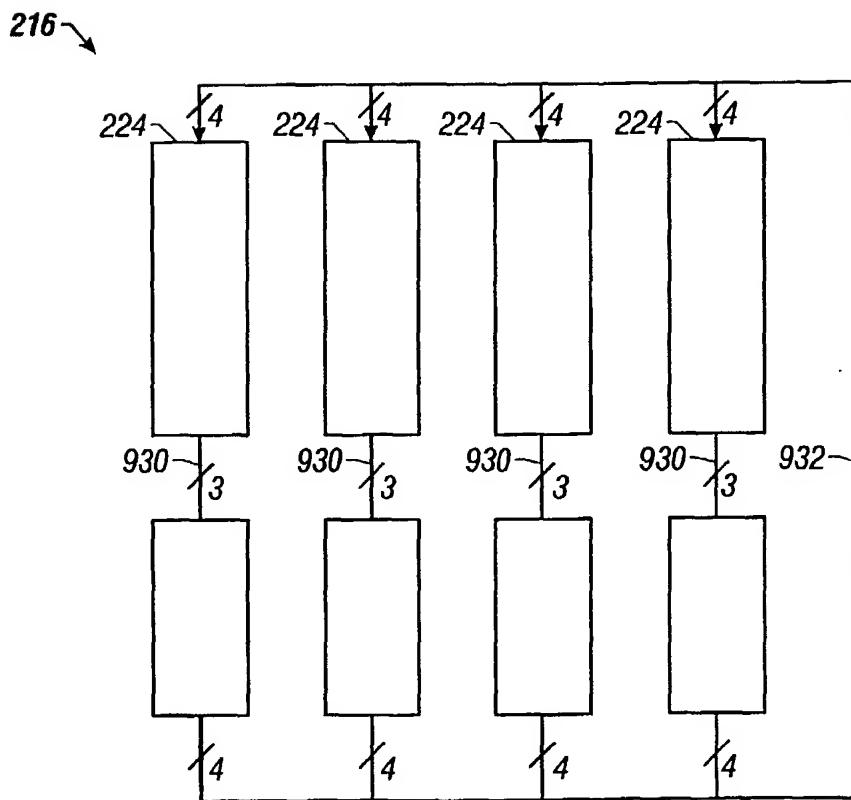


FIG. 9

7/7

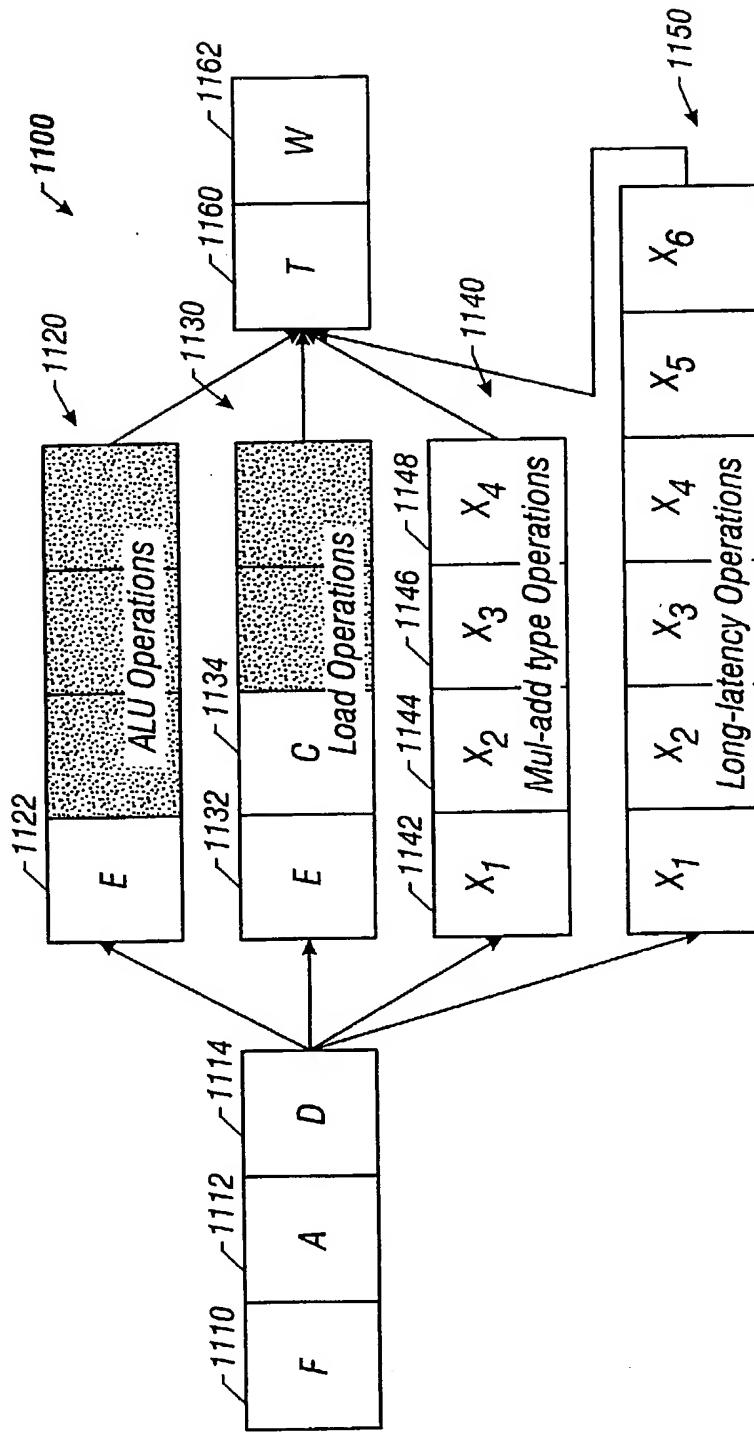


FIG. 10

INTERNATIONAL SEARCH REPORT

Int'l. Appl. No.
PCT/US 99/28820

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 G06F9/30

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the International search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category * | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|------------|---|-----------------------|
| X | EP 0 767 425 A (DIGITAL EQUIPMENT CORP) 9 April 1997 (1997-04-09) column 3, line 31 - line 42 column 6, line 8 - line 15 column 10, line 43 - line 48 column 11, line 19 - line 31 | 1,2,4-9, 16-21 |
| A | | 10,11, 14,22 |
| X | US 4 980 819 A (SHEN JIAN-KUO ET AL) 25 December 1990 (1990-12-25) column 2, line 28 - line 52 | 1,2,4-7, 16-19 |
| A | column 3, line 45 - line 53 column 4, line 54 - line 58 column 5, line 27 - line 60 column 7, line 10 - column 8, line 8 column 10, line 3 - line 20 figure 1 | 8,9,11, 14 |
| | | -/- |

Further documents are listed in the continuation of box C.

Patent family members are listed in annex.

* Special categories of cited documents :

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the International filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the International filing date but later than the priority date claimed

- "T" later document published after the International filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- "a" document member of the same patent family

Date of the actual completion of the International search

22 March 2000

Date of mailing of the International search report

30/03/2000

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl
Fax (+31-70) 340-3016

Authorized officer

Moraiti, M

INTERNATIONAL SEARCH REPORT

Intern. Appl. No.
PCT/US 99/28820

C/(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

| Category | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|----------|--|-----------------------|
| A | US 5 530 817 A (MASUBUCHI YOSHIO) 25 June 1996 (1996-06-25) column 3, line 1 - line 18 _____ | 1-3, 16 |
| A | EP 0 588 341 A (TOYOTA MOTOR CO LTD) 23 March 1994 (1994-03-23) page 3, line 54 -page 4, line 9 _____ | 1, 2, 7, 16, 19 |
| A | EP 0 676 691 A (HEWLETT PACKARD CO ; HITACHI LTD (JP)) 11 October 1995 (1995-10-11) column 5, line 25 -column 6, line 14 _____ | 10, 22 |

INTERNATIONAL SEARCH REPORT

Information on patent family members

Intell. Appl. No.

PCT/US 99/28820

| Patent document cited in search report | | Publication date | Patent family member(s) | | Publication date |
|--|---|------------------|-------------------------|------------|------------------|
| EP 0767425 | A | 09-04-1997 | NONE | | |
| US 4980819 | A | 25-12-1990 | NONE | | |
| US 5530817 | A | 25-06-1996 | JP | 5233281 A | 10-09-1993 |
| EP 0588341 | A | 23-03-1994 | JP | 6103068 A | 15-04-1994 |
| | | | DE | 69322566 D | 28-01-1999 |
| | | | DE | 69322566 T | 17-06-1999 |
| EP 0676691 | A | 11-10-1995 | JP | 7281897 A | 27-10-1995 |
| | | | US | 5564031 A | 08-10-1996 |